

Recenze projektu do předmětu PV070 – Digitální knihovny

Zpracoval:

Karel Dvořák

16. 1. 2006

1. Recenzovaný projekt:

AMeGA (Automatic Metadata Generation Application) Project

URL projektu: <http://ils.unc.edu/mrc/amega.htm>

Autorem projektu je Jane Greenberg a kolektiv, projekt byl konzultován se skupinou AMeGA Metadata Task Force [2].

Projekt byl řešen na universitě v Severní Karolině (University of North Carolina at Chapel Hill) a je součástí akčního plánu [4] Kongresové knihovny (Library of Congress).

Projekt je rozšířením projektu Metadata Generation Research [5].

2. Stručná charakteristika projektu:

Cílem projektu AMeGA bylo zjistit dosavadní stav na poli automatického generování metadat, zjistit, která metadata je vhodné získávat pomocí automatických nebo poloautomatických postupů a které postupy jsou nejvhodnější pro dané typy (digitálních) zdrojů a takto získané poznatky vydat jako doporučení pro tvůrce nástrojů na generování metadat.

3. Doba řešení, aktuální stav

Projekt byl řešen od začátku roku 2004 do února roku 2005.

Cíle projektu byly dosaženy. Hlavním výstupem projektu je závěrečná zpráva [3] obsahující (v kapitole 8) doporučenou funkcionalitu nástrojů pro automatické generování metadat ve verzi 1.0. Tvůrci předpokládají, že tato specifikace bude rozšířena a vylepšena komunitou pracovníků zabývajících se knihovnictvím a metadaty.

4. Cíle projektu

Hlavním cílem projektu AMeGA se bylo vytvořit vzorovou specifikaci pro vývoj nástrojů na automatické generování metadat, která byla vytyčena v akčním plánu [4] Kongresové knihovny, konkrétně v sekci 4.2. Dále bylo cílem projektu identifikovat a doporučit funkcionality pro aplikace podporující automatické generování metadat pro knihovní a bibliografickou komunitu.

Projekt AMeGA si stanovil tyto tři cíle:

1. Vyhodnotit úroveň funkcí pro automatické generování metadat u současných nástrojů pro vytváření obsahů a u nástrojů pro automatické vytváření metadat a přezkoumat, jakou funkcionalitu v tomto směru mají integrované knihovní systémy (integrated library systems – ILS).
2. Udělat průzkum mezi experty v oblasti metadat (profesionálními katalogizátory a indexátory

a osobami erudovanými v oblasti tvorby metadat) s cílem zjistit, které aspekty tvorby metadat jsou nejlépe automatizovatelné.

3. Vytvořit konečnou zprávu [3] s doporučenými funkcemi pro nástroje na automatické generování metadat.

Dalším výsledkem projektu bylo určení doporučených dalších kroků v této oblasti pro Kongresovou knihovnu.

5. Popis projektu a jeho výsledku

Podrobnější informace lze nalézt v závěrečné zprávě [3].

5.1 Řešení cíle 1 – Vyhodnocení současných nástrojů a postupů

- Bylo provedeno prozkoumání literatury zabývající se výzkumem v oblasti automatického generování metadat.
- Bylo proveden průzkum sedmi různých typů softwaru pro tvorbu obsahů zaměřený na podporovanou funkcionalitu v oblasti automatického generování metadat. Získané metadatové položky byly porovnány se standardem Dublin Core a významově podobné byly namapovány do elementů Dublin Core.

Shrnutí:

Výzkum v oblasti automatického generování metadat se dělí do dvou hlavních skupin: na *experimentální výzkum*, který se zaměřuje na techniky získávání informací a obsah digitálního zdroje, a *aplikovaný výzkum*, který se zaměřuje na vývoj software na vytváření obsahu a nástrojů na generování metadat.

Hlavním nálezem v této oblasti je, že zde dochází k odcizení experimentálního a aplikovaného výzkumu. Nástroje pro generování metadat by mohly být velmi zlepšeny integrováním výsledků experimentálního výzkumu.

Aplikace pro generování metadat by mohly být také vylepšeny, pokud by převzaly výhody funkcí pro generování metadat od software na vytváření obsahu.

5.2 Řešení cíle 2 – Průzkum mezi experty

Byl proveden průzkum s cílem identifikovat funkcionalitu požadovanou pro aplikace pro automatické generování metadat. Data byla sbírána z následujících zdrojů:

- od účastníků průzkumu, z jejich zkušeností v této oblasti
- ze současných postupů organizace metadat
- z názorů účastníků na automatické generování metadat pro standard Dublin Core
- z názorů účastníků na generování metadat obecně
- z názorů účastníků na požadované funkce aplikací na automatické generování metadat

Studie metadatových postupů používaných účastníky (i organizací) byla omezena na digitální objekty podobné dokumentům (digital document-like objects - DDLO) definované jako „primárně textové zdroje, které jsou přístupné přes webový prohlížeč“ [6].

Shrnutí:

- Bylo dotazováno 217 účastníků průzkumu (cílem bylo získat nejméně 100 účastníků).

- Tři čtvrtiny účastníků průzkumu měly tři a víceletou zkušenost s katalogizací a/nebo indexováním, což opravňuje označovat je jako experty v oblasti metadat.
- Organizace používají celou paletu různých metadatových standardů (například Machine Readable Cataloging (MARC), Dublin Core, Encoded Archival Description a další).
- Většina účastníků (81%) používá jeden nebo dva systémy pro tvorbu metadat, maximum bylo využití sedmi různých systémů.
- Hodnocení elementů Dublin Core
 - Účastníci předvídají větší přesnost při použití automatického zpracování pro technická metadata (např. ID, jazyk, formát) než pro metadata vyžadující rozumové zpracování (např. předmět a popis).
 - Z názorů účastníků na alokaci zdrojů pro automatické generování elementů Dublin Core plyne zásadní rozpor mezi užitečností a uskutečnitelností; účastníci zdůrazňují, že je potřebný další výzkum v oblasti pochopení metadat a procesů jejich tvorby.
- Účastníci podporují automatické generování metadat, ale většina (96%) z nich nedoporučuje plně automatické generování metadat, spíše preferují nejprve použít automatických prostředků a které následně umožní manuální změny a kontrolu.
- Účastníci naznačují, že velmi důležitá, a v některých případech i kritická, je podpora automatického generování metadat pro netextové zdroje.

5.3 Řešení cíle 3 - Doporučené funkce pro nástroje na automatické generování metadat

Byla vytvořena studie s doporučenou funkcionalitou pro aplikace na automatické generování metadat. Tato studie je součástí závěrečné zprávy celého projektu [3], kapitola 8.

Obsahuje shrnutí a doporučené postupy a funkce získané při řešení předchozích cílů. Je členěna do těchto okruhů:

- Cíle systému
- Obecná doporučení
- Konfigurace systému
- Identifikace a sběr metadat
- Podpora lidské práce při generování metadat
- Vylepšení a publikování metadat
- Vyhodnocení kvality získaných metadat
- Generování metadat pro netextové zdroje

Dalším výsledkem projektu je zpráva s doporučeními dalších kroků pro Kongresovou knihovnu, zpráva je součástí závěrečné zprávy projektu [3] – kapitola 10.

Obsahuje tři hlavní doporučení:

- vytvořit aplikaci pro automatické generování metadat
- podporovat a rozvíjet výzkum automatického generování metadat
- vytvořit mechanismus pro komunikaci a vyjednávání s ostatními dodavateli metadatového software s cílem zlepšit funkčnost těchto aplikací v oblasti metadat

6. Vlastní zhodnocení projektu a jeho přínosu

Projekt AMeGa dle mého názoru nepřinesl žádné převratné myšlenky – velká část věcí mi přišla známá a logická (může to být způsobeno absolvováním přednášek z předmětu Digitální knihovny), nicméně projekt považuji za prospěšný z těchto dvou důvodů:

- pokusil se systematicky shrnout dosavadní roztrášené poznatky a postupy
- vytyčil funkce, které by měl podporovat každý nástroj pro tvorbu metadat, což by mělo vyústit v kompatibilitu získaných metadat a tím i napomoci jejich většímu využití

7. Seznam zdrojů

- [1] Web projektu *AMeGA*: <http://ils.unc.edu/mrc/amega.htm>
- [2] Web pracovní skupiny *AMeGA Metadata Task Force*: http://ils.unc.edu/mrc/amega_task.htm
- [3] Závěrečná zpráva projektu *AmeGA*:
http://www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf
- [4] Akční plán kongresové knihovny: <http://www.loc.gov/catdir/bibcontrol/actionplan.pdf>
- [5] Projekt *Metadata Generation Research*: http://ils.unc.edu/mrc/mgr_index.htm
- [6] Definice pojmů použité v průzkumu: http://ils.unc.edu/mrc/amega_survey_defs.htm

8. Metadata recenze ve standardu Dublin Core

Dublin Core atribut s kvalifikátorem	Schéma	Hodnota
Title		Recenze projektu AMeGA do předmětu PV070 – Digitální knihovny
Creator		Karel Dvořák
Subject		AMeGa
Subject		metadata
Subject		generování
Subject		automatické
Description.abstract		Recenze projektu AMeGA (Automatic Metadata Generation Application project). Projekt AMeGA se zabývá vytvořením doporučené specifikace nástrojů pro automatickou tvorbu metadat.
Date.created		16. 1. 2006
Type	DCMIType	Text
Format	IMT	application/pdf
Format.medium		computerFile
Format.extent		5 stran formátu A4
Identifier		http://www.fi.muni.cz/~xdvora15/PV070/dvorak_amega.pdf

Source	URL	http://ils.unc.edu/mrc/amega.htm
Language	RFC3066	cze